

UESTC at ImageCLEF 2010 medical retrieval task

Hong Wu, Changjun Hu, Sikun Chen

School of Computer Science and Engineering
University of Electronic Science and Technology of China
Chengdu 611731, P. R. China
hwu@uestc.edu.cn

Abstract. This paper presents the UESTC contribution to the ImageCLEF 2010 medical retrieval task. For ad-hoc retrieval and case-based retrieval, we only use text information, and propose a phrase-based approach. Phrases, subphrases and individual words are used with vector space model (VSM) for ranking. Phrases and subphrases are extracted with the help of MetaMap, and all extracted phrasal terms are corresponding to concepts in UMLS. Two term weighting methods are proposed, one is to weight terms with their idfs, and the other is adapted to assign lower weights to phrasal terms. We also propose a query expansion method which can extract more phrases for query by relaxing the restrictions on phrase extraction. For modality classification, we use three global texture features with SVM and Ada-boost.MH respectively.

Keywords: text retrieval, image retrieval, medical retrieval, modality classification, phrase extraction, MetaMap, UMLS

1 Introduction

This paper describes the first participation of the School of Computer Science and Engineering at University of Electronic Science and Technology of China (UESTC) in the ImageCLEF 2010 medical retrieval task.

ImageCLEFmed'10 [1] includes three types of tasks, ad-hoc retrieval, case-based retrieval and modality classification. For the retrieval tasks, the dataset similar to 2008 and 2009 is used but with a larger number of images. The dataset contains all images (>77,000) from articles published in Radiology and Radiographics including the text of the captions and a link to the html of the full text articles. In the ad-hoc retrieval task, a set of textual queries, each of which with several sample images, are given, and the goal is to retrieve the images most relevant to each topic. In the case-based task, a set of case-based information requests are given, and the goal is to retrieve the articles most relevant to the topic case. In the modality classification task, training and testing medical images are given for classification based on their modality, such as CT, MR, XR etc.

In this paper, we describe our phrase-based approach to two retrieval tasks and classification algorithm for modality classification. For retrieval tasks, only text information of title and caption is used. Phrases, subphrases and individual words

are used as indexing terms with vector space model (VSM). Phrases and subphrases are extracted with the help of MetaMap¹, so that all the phrasal terms are corresponding to concepts in UMLS². Since the text information for ad-hoc is very short, it is necessary to adapt traditional term weighting methods. We propose to weight terms with their *idfs*, and measure the similarity with dot-product. But in this way, the phrasal terms are always over-rewarded. Then we give another weighting method which assigning lower weights to phrasal terms. We also propose a query expansion method which can extract more phrases for query by relaxing the restrictions on phrase extraction. For modality classification, we use three global texture features with SVM [2] and Ada-boost.MH [3] respectively.

The remainder of this paper is organized as follows. The phrase-based retrieval approach and modality classification algorithm are described in section 2 and 3 respectively. And our submitted runs and results are presented in section 4, followed by the conclusions and future works in section 5.

2 Phrase-based Medical Retrieval

2.1 Using phrase as indexing term

The selection of appropriate indexing terms is critical to information retrieval. Traditional retrieval systems use word or word stem as indexing term. And these representations of content are usually inadequate since single words are rarely specific enough for accurate discrimination. A better method is to identify groups of words that form meaningful phrases, especially if these phrases denote important concepts in the related domain. This is corresponding to using phrase or concept as indexing term. In the past years, concept-based approaches have been investigated in ImageCLEFmed [4, 5, 6], but to the best of our knowledge, there's still no work using phrase as indexing term in this campaign. In this year, we investigate the phrase-based medical retrieval.

In the past, various types of phrases, such as sequential n-grams [7], head-modifier pairs extracted from syntactic structures [8, 9, 10, 11], proximity-based phrases [12], were examined with conventional retrieval models (e.g. vector space model). In our approach, we consider phrases which are corresponding to medical concepts. The phrases are extracted with the help of MetaMap, which is a highly configurable program to map biomedical text to the UMLS Metathesaurus. MetaMap maps the longest possible phrase to concept so that it discovers the most specific concept possible. This may cause much mismatch between query terms and document terms, if the detected concepts (CUIs) are directly used as indexing term. Because a general concept and a specific concept may all be relevant to a user's need, also the meaning of a concept can be expressed by a phrase or several words or phrases co-occur in the context. [13] has given an example for this:

¹ <http://mmtx.nlm.nih.gov/>

² <http://www.nlm.nih.gov/research/umls>

The 24th query of ImagCLEF2005 is “Show me images of right middle lobe pneumonia”, and the best mapping schema of Metamap will give these concepts:

“C0150627” (Images)

“C0578577” (Right middle lobe pneumonia)

But the relevant documents contain concepts, “C0032285” (pneumonia) or “C0796494” (lobe) will not match with query concepts, and then will get an unfavorable ranking.

One way to tackle this problem is to expand query or document with some concepts related to the mapped concepts, e.g. hypernyms or hyponym [13]. Following this way, we give the derivation of our approach.

When mapping a phrase to concept, MetaMap also generates some candidates. Candidates consist of one or more constituent words or their variants of the phrase, and are corresponding to concepts in UMLS. Some of these concepts are related to the mapped most specific concept, and can be used to expand query or document. But when generating candidate, a phrase can be mapped to several concepts (more frequently for subphrase), and much noise will be introduced if all corresponding concepts are added. So, we consider using phrase (subphrase) instead of concept (CUI) to represent document, and phrases, subphrases and individual words are all used as index terms. The subphrases of a noun phrase capture a part of the meaning of the noun phrase, and can be regarded as a weak representation of its meaning. And the use of both phrase and its subphrases can increase the chance of match between query and document having difference linguistic forms of similar meaning.

2.2 Phrase Extraction

In our experiments, we do not develop a phrase extraction algorithm. All phrases and their subphrase are extracted by MetaMap.

MetaMap [14] performs the following steps to map text to concept for each textual utterance:

1. *Parse the text into noun phrases and perform the remaining steps for each phrase;*
2. *Generate the variants for the noun phrase where a variant essentially consists of one or more noun phrase words together with all of its spelling variants, abbreviations, acronyms, synonyms, inflectional and derivational variants, and meaningful combinations of these;*
3. *Form the candidate set of all Metathesaurus strings containing one of the variants;*
4. *For each candidate, compute the mapping from the noun phrase and calculate the strength of the mapping using an evaluation function. Order the candidates by mapping strength; and*
5. *Combine candidates involved with disjoint parts of the noun phrase, recompute the match strength based on the combined candidates, and select those having the highest score to form a set of best Metathesaurus mappings for the original noun phrase.*

The best candidate is corresponding to the longest phrase, and other candidates corresponding to its subphrases or constituent words. The phrases and multi-word subphrases are added to query and document before indexing and retrieval.

MetaMap is designed for mapping the longest possible phrase to concept, not for phrase and subphrase extraction. We find that it is not easy to control it for phrase extraction. For example, When processing 8th query: “microscopic images of streptococcus pneumonia”, MetaMap generates a candidate “streptococcus pneumoniae” with LexVariation=0.5 due to an inflectional variation. For “chest x-ray” in 17th query, MetaMap generates an unwanted candidate “breast x-ray” with LexVariation=2. There’s no simple rule to select candidates having meaning closely related to the best candidate, and we use a strict rule.

In experiments, we use 0910 Strict Model Dataset for MetaMap. When calling MetaMap for phrase extraction, we do not allow derivational variants by setting the parameter ‘-d’, because derivational variants always involve a significant change in meaning. From output files of MetaMap, only candidates with ‘LexVariation’=0 and ‘MatchedWords Count’>1 are selected to form the phrasal terms. ‘LexVariation’=0 implies that no lexical variances are permitted in phrase extraction, and ‘MatchedWords Count’>1 means only multi-words phrases are selected.

2.3 Term weighting

We use phrasal terms and single word terms with VSM, and propose two term weighting methods. For the ad-hoc retrieval task, the context information (title and caption) is much shorter than document in traditional IR. Thus we think that the term frequency (tf) may be not important in this case, and use a simple term weighting method, where only idfs of indexing terms are used for term weighting. This term weighting is also used for case-based retrieval. And the similarity between query and document is measured by dot product of query vector and document vector.

When using VSM to combine weights of phrases, subphrases and single word terms, phrasal terms are over-rewarded. Since occurrence of a phrase in a document also indicates the occurrence of its subphrases and constituent words. To solve this problem, we propose another term weighting method which associates lower weights to phrasal terms. For convenience of description, we introduce some concepts to describe the relationship between phrases. We say a phrase or single word A is an *offspring component* of a phrase B , if and only if it is a subphrase or constituent word of phrase B . We say a phrase or single word A is a *son component* of a phrase B , if and only if A is an *offspring component* of B , and there’s no *offspring component* of B which having A as its *offspring component*. In the second term weighting method, the weight of phrasal term is changed to be its *idf* minus the maximum *idf* of its *son components*.

2.4. Query Expansion

Our query expansion algorithm is just to relax the restrictions on phrase extraction. Candidates with ‘MatchedWords Count’>1 (multi-words) are selected to form the phrasal terms for query, and more phrase terms are extracted than previous setting.

3. modality classification

For modality classification, we use three global texture features: LBP texture feature, Gabor texture feature and Tamura texture feature.

LBP: Local Binary Pattern (LBP) [15] features have performed very well in various applications, including texture classification and segmentation, image retrieval and surface inspection. In our experiments, LBP operator with 8 neighbors on a circle of radius 4 is applied to each pixel, and the obtained results are cumulated to form 256-dim LBP histogram.

Tamura Texture Feature: Based on the research of textural features corresponding to human visual perception, Tamura et al.[16] proposed six basic textural features, namely, coarseness, contrast, directionality, line likeness, regularity, and roughness. In our experiments, coarseness, contrast and directionality features are computed on a per-pixel basis, and the values are quantized into a three-dimensional histogram ($8 \times 8 \times 8=512$ bins) to form one 512-dim vector.

Gabor Texture Feature: Gabor filter based approaches are popular for texture feature extraction. Based on the work of Manjunath et al [17], gabor filters with 3 scales and 4 orientations are used to filter image, and the values in the filtered images are quantized to 10 bins to form a 120-dim histogram feature.

For feature combination, they are simply concatenated to form an 888-dim feature vector.

We use two algorithms for classification. One is SVM [3] with rbf kernel, and one-vs-one strategy is used for multi-class classification. LibSVM [18] is used in our experiments, and the parameters are tuned by cross-validation on training data. The other is Adaboost.MH [4], a multi-class boosting algorithm. An implementation named MultiBoost [19] is used in our experiments.

4. Submitted Runs and Results

For ad-hoc retrieval, collection with title and caption only is used since it was proven to be effective and obtain the best results in ImageCLEFmed 2008 [20]. After phrasal terms were added, collection is indexed by Lemur IR toolkit³. We also update the stop word list to add common terms found in the queries that are not relevant to medical domain such as ‘image’, ‘photo’, and ‘figure’. For convenience, this procedure has also been applied to case-based retrieval without update. But the use of title and caption only may lose important information for case-based retrieval and result in poor performance.

4.1 Ad-hoc Retrieval

We have submitted the following 3 textual runs for the 16 ad-hoc topics [1]:

³ <http://www.lemurproject.org/>

(1) *UESTC_image_pBasic*: Phrasal terms are extracted by the approach described in Section 2.2, and terms are weighted by their *idfs* as the first method in Section 2.3. Similarity is measured by dot product of query vector and document vector.

(2) *UESTC_image_pNw*: It is similar to the above basic run (*UESTC_image_pBasic*), but the term weighting method is changed to the second method in Section 2.3.

(3) *UESTC_image_pQE*: This run is similar to the basic run (*UESTC_image_pBasic*), but a query expansion method (Section 2.4) is used to get more phrasal terms for query.

To evaluate the effectiveness of the use of phrasal term, we conduct an additional run *Image_word_idf* when preparing this report. *Image_word_idf* uses word stem as indexing term, and the term weighting is the same as *UESTC_image_pBasic*.

Table1. Results of the Submitted and additional Runs for the Ad-hoc Retrieval

Runs	MAP	bPref	P10
<i>UESTC_image_pQE</i>	0.2789	0.297	0.3125
<i>UESTC_image_pNw</i>	0.2751	0.3028	0.3438
<i>UESTC_image_pBasic</i>	0.2713	0.2963	0.3438
<i>Image_word_idf</i>	0.2216	0.2499	0.3375

Table 1 gives the results of our three submitted runs and the additional run for the ad-hoc retrieval. The performances of the three submitted runs are very similar. The performance of *UESTC_image_pNw* is better than *UESTC_image_pBasic*, but the improvement is subtle. *UESTC_image_pQE* achieves the best MAP (0.2789) of our submitted runs, and is ranked 3rd among all best official runs of each group for automatic textual retrieval. But P10 of *UESTC_image_pQE* is lower than the other two runs. The MAPs and bPrefs of the three phrase-based approaches are apparently superior to word stem based approach *Image_word_idf*. This may hint the use of phrase in medical retrieval. Table 2 presents the performance of best official runs of each group for automatic textual retrieval. The third run with bold text is our best official textual run.

Table 2. Performance of Best Official Runs of Each Group for Automatic Textual Retrieval

Runs	MAP
<i>ipl_aueb_AdHoc_default_TC</i>	0.3235
<i>OHSU_pm_all_all_mod</i>	0.3029
<i>UESTC_image_pQE</i>	0.2789
<i>issr_CT</i>	0.2583
<i>runfile_hes-so-vs_image-based_captions</i>	0.2568
<i>ad_hoc_QE_0.1_Citations_and_All_Image_Text</i>	0.188
<i>NMFText_k2_11</i>	0.1005

Originally with our method, the 16th ad-hoc topic “images of dermatofibroma” will match no document. And we update this query by inserting a space character to “images of dermato fibroma”, when conducting the submitted runs. From the raw results, we calculate the corrected results which corresponding to runs without update to 16th topic. The corrected MAPs of *UESTC_image_pQE*, *UESTC_image_pNW*, *UESTC_image_pBasic* are 0.2777, 0.2739, and 0.2701 respectively. The corrected bPrefs of them are 0.2969, 0.3027, and 0.2962 respectively. And the P10s of them stay unchanged. The differences between the corrected and original results are subtle, and will not affect the conclusions on the results of submitted runs.

4.2 Case-based Retrieval

The methods used in ad-hoc retrieval are directly used for case-based retrieval, and 3 textual runs are submitted for 14 case-based topics [1].

- (1) *UESTC_case_pBasic*: This run uses the same method as *UESTC_image_pBasic*.
- (2) *UESTC_case_pNw*: This run uses the same method as *UESTC_image_pNw*.
- (3) *UESTC_case_pQE*: This run uses the same method as *UESTC_image_pQE*.

For evaluation, we also conduct an additional run *Case_word_idf*, and the methods used in them are the same as *Image_word_idf*. All the four runs are automatic textual runs.

Table 3. Results of the Submitted Runs and additional Runs for the Case-based Retrieval

Runs	MAP	bPref	P10
<i>UESTC_case_pBasic</i>	0.1692	0.184	0.2643
<i>UESTC_case_pQE</i>	0.1677	0.1852	0.2786
<i>UESTC_case_pNw</i>	0.1522	0.1725	0.2714
<i>Case_word_idf</i>	0.1501	0.1688	0.2786

Table 3 gives the results of our three submitted runs and the additional run for the case-based retrieval. *UESTC_case_pBasic* achieves the best MAP (0.1692) of our submitted runs, and is ranked 6th among all 7 groups who submitted for automatic textual runs. The best MAP is 0.2902, achieved by UIUCIBM. The low performance of our approach in case-based retrieval may have two reasons, the first is that only article title and image caption are used for indexing, some important information may lose in the process; the second is that our methods originally proposed for ad-hoc retrieval may not be suitable for case-based retrieval.

4.3 Modality Classification

We use LBP texture feature, Gabor texture feature and Tamura texture feature for modality classification, and submit two visual runs with different classifier.

- (1) *UESTC_modality_boosting*: This run uses Adaboost.MH with the three global texture features for modality classification.

(2) *UESTC_modality_svm*: This run uses SVM with the three global features for modality classification.

Table 4. Results of the Submitted Runs for Modality Classification

Run	Accuracy
<i>UESTC_modality_boosting</i>	0.82
<i>UESTC_modality_svm</i>	0.80

Table 4 presents the classification accuracies of the two submitted runs. *UESTC_modality_boosting* achieves the best accuracy of 0.82, and is ranked 2th among all 5 groups who submitted visual runs.

5. Conclusions and Future Work

This paper describes our contribution to the ImageCLEF 2010 medical retrieval task. For ad-hoc retrieval, we have submitted 3 runs with our phrase-based approaches. With the same methods, 3 runs have been submitted for case-based retrieval. For modality classification, 2 runs have been submitted, using global texture features with two different classifiers respectively. The runs submitted to ad-hoc retrieval and modality classification are successful, and achieve 3th rank in automatic textual retrieval and 2th in modality classification.

Our research on medical retrieval is still primary, both the phrase extraction and term weighting. And there's no extensive comparison of different methods. In the future, we will develop and compare different phrase extraction algorithms and term weighting schemes, and use more text features for case-based retrieval. For modality classification, we plan to test other visual features and advanced classification algorithms.

Acknowledgments. This research is partly supported by the National Science Foundation of China under grants 60873185 and by the Key Program of the Youth Science Foundation of UESTC under Grant JX0745.

References

1. Henning Müller, Jayashree Kalpathy-Cramer, Ivan Eggel, Steven Bedrick, Charles E. Kahn Jr., and William Hersh. Overview of the CLEF 2010 medical image retrieval track. In the Working Notes of CLEF 2010, Padova, Italy, (2010)
2. Vapnik, V.N.: The nature of statistical learning theory. Springer, Heidelberg (1995)
3. Schapire, R. & Singer, Y., 'Improved boosting algorithms using confidence-rated prediction', *Machine Learning* 37(3), 297–336, (1999)
4. Lacoste, C., Chevallet, J.P., Lim, J.H., Wei, X., Raccoceanu, D., Le, T.H.D., Teodorescu, R., Vuillenemot, N.: Ipal knowledge-based medical image retrieval in imageclefmed 2006. In: Working Notes for the CLEF 2006 Workshop, Alicante, Spain, September 20-22 (2006)

5. Chevallet, J.P., Lim, J.H., Le, T.H.D.: Domain knowledge conceptual inter-media indexing, application to multilingual multimedia medical reports. In: ACM Sixteenth Conference on Information and Knowledge Management (CIKM 2007), November 6-9 (2007)
6. Maisonnasse, L., Gaussier, E., Chevallet, J.P.: Multiplying concept sources for graph modeling. In: Peters, C., Jijkoun, V., Mandl, T., Müller, H., Oard, D.W., Peñas, A., Petras, V., Santos, D. (eds.) CLEF 2007. LNCS, vol. 5152. Springer, Heidelberg (2008)
7. Mitra, M., Buckley, C., Singhal, A., and Cardie, C. An analysis of statistical and syntactic phrases. In Proceedings of RIAO '97, pages 200–214, (1997)
8. Lewis, D.D., and Croft., W.B., Term clustering of syntactic phrases. In Proceedings of SIGIR '90, pages 385–404, (1990)
9. Zhai, C., Fast statistical parsing of noun phrases for document indexing. In Proceedings of ANLP '97, pages 312–319, (1997)
10. Dillon, M., and Gray, A.S. Fasit: A fully automatic syntactically based indexing system. *Journal of the American Society for Information Science*, 34(2):99–108, (1983)
11. Strzalkowski, T., Perez-Carballo, J., and Marinescu, M. Natural language information retrieval: Trec-3 report. In Proceedings of TREC-3, pages 39–54, (1994)
12. Turpin, A., and Moffat, A., Statistical phrases for vector-space information retrieval. In Proceedings of SIGIR '99, pages 309–310, (1999)
13. Le T.H.D., Chevallet, J.-P., Dong T.B.T., Thesaurus-based query and document expansion in conceptual indexing with UMLS: Application in medical information retrieval, *IEEE International Conference on In Research, Innovation and Vision for the Future*, pp. 242-246, (2007)
14. Aronson, A. R., *MetaMap: Mapping Text to the UMLS Metathesaurus*, <http://skr.nlm.nih.gov/papers/references/metamap06.pdf> , July (2006)
15. Ojala, T., Peitikkäinen, M., and Mäenpää, T., “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971–987, July (2002)
16. Tamura, H., Mori, S., and Yamawaki, T., Texture features corresponding to visual perception. *IEEE Trans. On Systems, Man, and Cybernetics*, 8(6) , (1978)
17. Manjunath, B., and Ma, W., Textures for browsing and retrieval of image data. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 18(8):837-842, (1996)
18. Chang, C.C. and Lin, C.J., *LIBSVM: a library for support vector machines*, (2001)
19. Fekete, R.B., Casagrande, N., Keggl, B.,: *MultiBoost*: <http://mloss.org/software/view/246/>
20. García-Cumbreras, M.A., Díaz-Galiano, M.C., Martín-Valdivia, M.T., Ureña López, L.A.: *SINAI at ImageCLEFphoto 2008*. In: *On-line Working Notes, CLEF 2008* (2008).